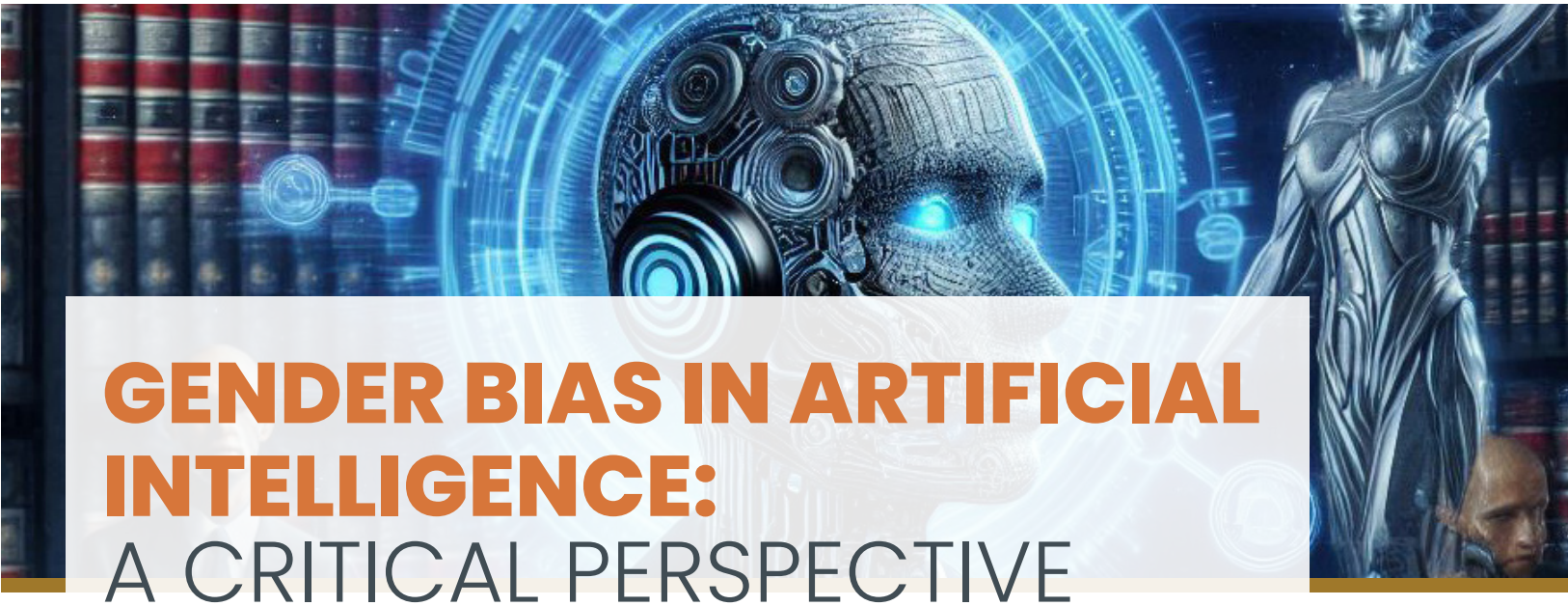


GENDER BIAS IN ARTIFICIAL INTELLIGENCE: A CRITICAL PERSPECTIVE AND LEGAL ANALYSIS

Dra. Trilce Fabiola Ovilla Bueno

Mexican lawyer. PhD in Law. She holds a master's degree in Corporate Law and has master's studies in Constitutional Procedural Law. She obtained a specialized degree in Political Training: Europe-Latin America Perspectives. Full-time professor at the School of Law at UNAM.

Abstract: Artificial Intelligence (AI) is transforming key industries like employment, healthcare, and criminal justice, but it also introduces significant ethical and legal challenges, particularly regarding gender bias. AI systems, often trained on biased historical data, can perpetuate and even amplify existing gender inequalities. This essay examines the legal implications of gender bias in AI, focusing on challenges to anti-discrimination laws, transparency issues, and the need for regulatory oversight. Gender bias in AI arises when systems are trained on datasets that reflect societal inequalities, leading to discriminatory outcomes. This bias is not a technical flaw, but rather a consequence of using data and algorithms that mirror patterns of discrimination. The lack of diversity among AI developers, who are predominantly male, exacerbates this issue by failing to account for the perspectives and needs of women and marginalized groups.

In legal contexts, the use of AI in hiring, criminal justice, and risk assessment raises ethical concerns. AI-driven systems risk reinforcing historical gender biases, which can undermine fairness in decision-making processes. Unchecked, these biases could worsen disparities in critical areas such as recruitment and justice administration, threatening legal protections. A major legal challenge is how AI interacts with existing anti-discrimination laws. To address these challenges, transparency in AI decision-making is

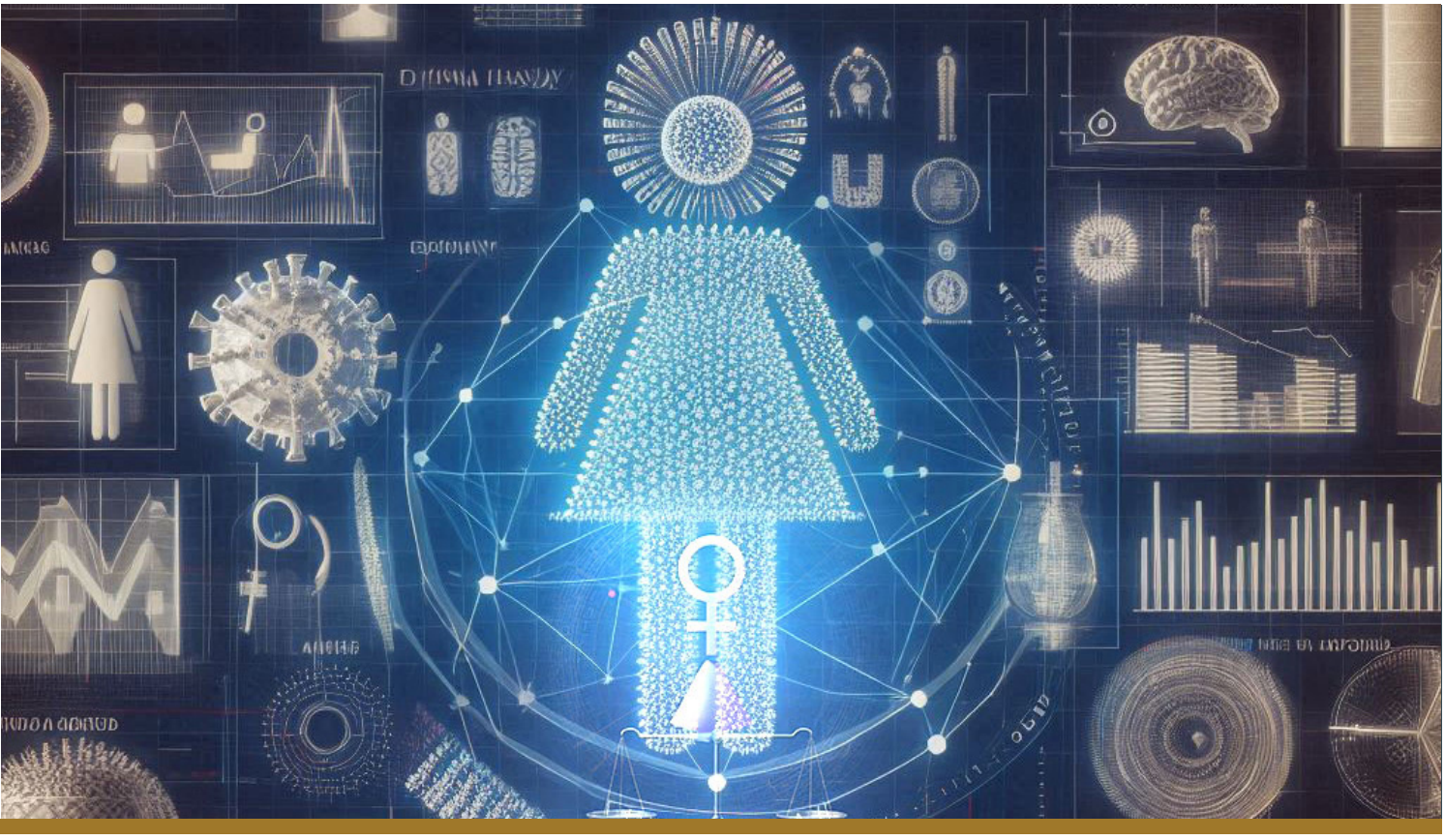
essential. Regulatory frameworks must evolve to require regular audits of AI systems and enforce accountability for biased outcomes. Ethical guidelines are insufficient; mandatory legal oversight is needed to ensure AI promotes fairness and inclusivity.

Keywords: Artificial Intelligence, gender bias, ethical, women rights.

Resumen: La inteligencia artificial (IA) está transformando industrias clave, como el empleo, la atención médica y la justicia penal, pero también plantea importantes desafíos éticos y legales, en particular en relación con el sesgo de género. Los sistemas de IA, a menudo entrenados con datos históricos con sesgos, pueden perpetuar e incluso amplificar las desigualdades de género existentes. Este artículo aborda las implicaciones legales del sesgo de género en la IA, centrándose en los desafíos actuales para las leyes contra la discriminación, los problemas de transparencia y la necesidad de supervisión regulatoria. El sesgo de género en la IA surge cuando los sistemas se entrenan con conjuntos de datos que reflejan desigualdades sociales, lo que lleva a resultados discriminatorios. Este sesgo no es un defecto técnico, sino una consecuencia del uso de datos y algoritmos que replican patrones de discriminación. La falta de diversidad entre los desarrolladores de IA, que son predominantemente hombres, agrava este problema al no considerar las perspectivas y necesidades de las mujeres y los grupos marginados.

En contextos legales, el uso de la IA en la contratación, la justicia penal y la evaluación de riesgos plantea preocupaciones éticas. Los sistemas impulsados por IA corren el riesgo de reforzar los sesgos de género históricos, lo que puede socavar la equidad en los procesos de toma de decisiones. Si no se corrigen, estos sesgos podrían agravar las disparidades en áreas críticas como la contratación y la administración de justicia, amenazando las protecciones legales. Un gran desafío legal es cómo interactúa la IA con las leyes sobre antidiscriminación existentes. Para abordar estos desafíos, la transparencia en la toma de decisiones de la IA es esencial. Los marcos regulatorios deben evolucionar para exigir auditorías regulares de los sistemas de IA y hacer cumplir la responsabilidad por los resultados sesgados. Las pautas éticas actuales son insuficientes para controlar esto; se necesita supervisión legal obligatoria para garantizar que la IA promueva la equidad y la inclusión.

Palabras clave: Inteligencia artificial, sesgo de género, ética, derechos de las mujeres.



Summary: I. Introduction; II. Sources of Gender Bias in AI (Data and Algorithmic Bias); III. Gender Bias in AI and the Law; III.1. Anti-Discrimination Law; III.2. Transparency and Accountability in AI Systems; IV. The Role of Regulatory Oversight and Ethical AI; V. Conclusion; VI. References.

I. Introduction

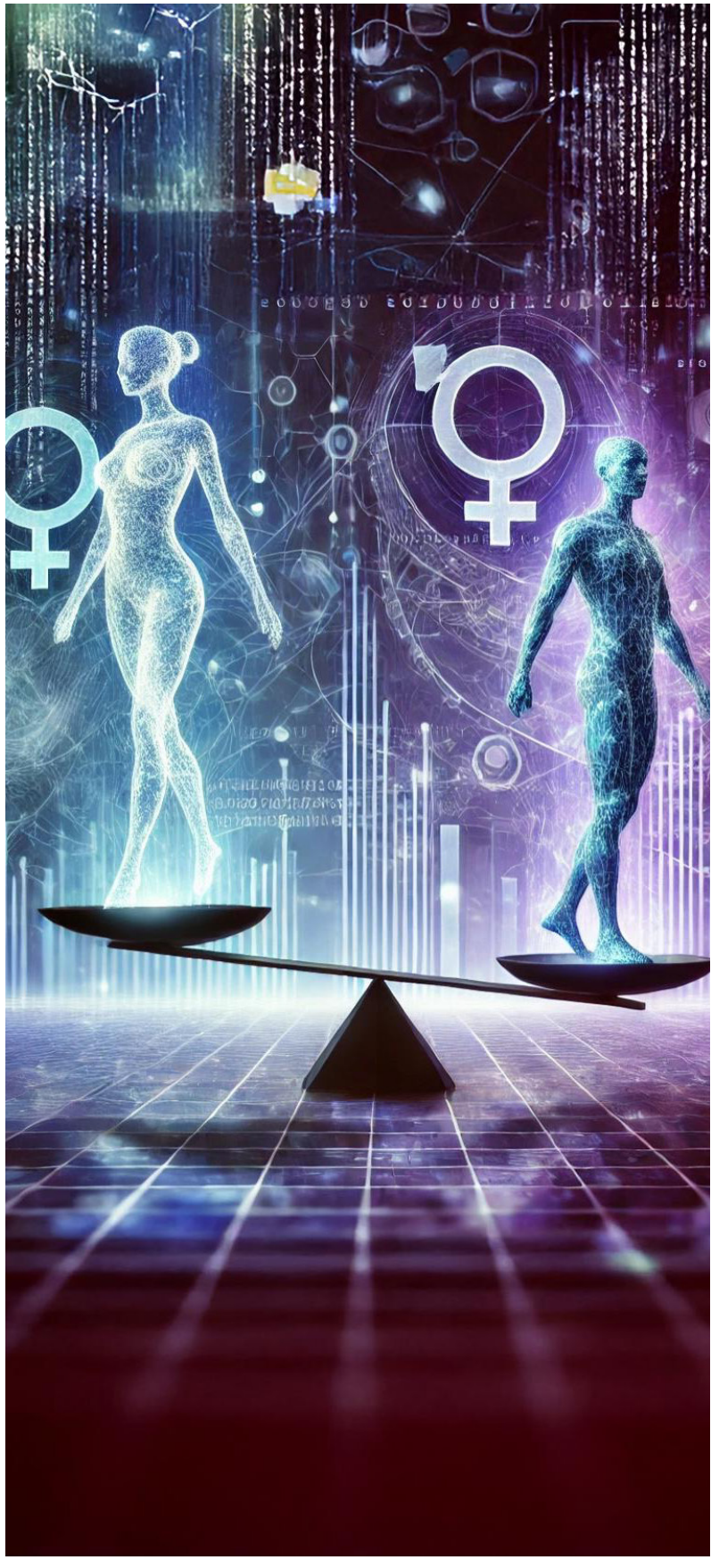
Artificial Intelligence (AI) is rapidly transforming modern society, revolutionizing industries and changing the way decisions are made across a wide array of sectors. From hiring practices to healthcare diagnostics, AI plays a critical role in shaping outcomes that directly impact individuals and communities. Yet, alongside its promises, AI raises significant ethical and legal concerns, particularly regarding gender bias. AI systems have the capacity to perpetuate and even exacerbate existing gender inequalities in various areas, such as employment, healthcare, and criminal justice. This article explores the issue of gender bias in AI from a legal perspective, focusing on how anti-discrimination laws are being challenged by AI technologies, the need for transparency, and the role of regulatory frameworks in addressing these issues.

At its core, gender bias in AI refers to situations where AI systems display prejudicial behavior based on gender. This bias often emerges when the AI learns from historical data that reflect the inequalities and stereotypes embedded in society. The data used to train AI models are typically rich with patterns that may inadvertently favor one gender over another, perpetuating existing societal biases.

The problem of gender bias in AI is not a design flaw per se but an inherent consequence of training AI with biased data and using algorithms that may not be well thought out. While the technology itself is neutral, the systems it creates can replicate and even amplify inequalities if they are not designed and implemented with care. Feminist theory, particularly feminist technoscience, offers valuable insight into the relationship between gender and technology. Scholars such as Sandra Harding and Donna Haraway have argued that technology is shaped by the cultural and societal forces in which it is developed, and AI is no exception.

The lack of diversity among AI developers is a key reason gender bias persists in these systems. The predominance of male developers in the tech industry leads to a narrow perspective that often overlooks the experiences and needs of women and marginalized groups. This is not necessarily intentional but reflects the blind spots that emerge when diverse viewpoints are not considered.

In her book Automating Inequality (2018), Virginia Eubanks illustrates how AI systems deployed in public services often perpetuate the very inequalities they aim to address. Eubanks demonstrates how automated systems can disproportionately harm women, especially those from low-income backgrounds, by replicating biases rooted in social and economic structures.



II. Sources of Gender Bias in AI (Data and Algorithmic Bias)

One of the most significant sources of gender bias in AI is biased data. AI systems depend on large datasets to make predictions and decisions. If the data reflects past discrimination or gender imbalances, the AI will likely reproduce those biases. For instance, women —particularly women of color— are often underrepresented in the datasets used to train AI models. In facial recognition systems, for example, in research by Joy Buolamwini and Timnit Gebru,[1] it was found that error rates were significantly higher when identifying dark-skinned women compared to white men.

In addition to underrepresentation, historical inequalities are also embedded in AI. When AI is trained on data that reflects past injustices, it reinforces these inequalities. Digital assistants like Siri and Alexa, for example, are given female voices, reinforcing traditional gender roles by portraying women as helpers or caregivers. These choices, while seemingly innocuous, reflect and reinforce societal stereotypes about gender.

A well-known case of gender bias in AI is Amazon's AI recruitment tool, which was found to systematically discriminate against women applying for technical positions.[2] The AI system had been trained on resumes submitted over a decade, during which men dominated technical roles. As a result, the AI favored male candidates and penalized resumes that mentioned women's colleges or activities.

Even without overtly biased data, algorithms themselves can introduce gender bias. Because algorithms are designed by humans, they can reflect the conscious or unconscious biases of their creators. Another issue is the feedback loop: if an AI system starts with biased assumptions, those biases can be reinforced and become more entrenched over time.[3]

Moreover, AI systems do not develop in isolation; they are created within specific cultural contexts that shape their design. The tech industry, long criticized for its lack of diversity, is predominantly male. This lack of diversity can lead to blind spots in the design process, where the needs and experiences of women are insufficiently considered.

[1] Buolamwini, Joy y Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", en *Proceedings of the 1st Conference on Fairness, Accountability, and Transparency (PMLR)*, 2018 [en línea], <https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.

[2] Dastin, Jeffrey, "Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women", en *Reuters*, 10 de octubre, 2018 [en línea], <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.

[3] Noble, Safiya Umoja, *Algorithms of Oppression: How Search Engines Reinforce Racism*, [s.d.], New York University Press, 2018.

III. Gender Bias in AI and the Law

The growing use of artificial intelligence (AI) in the legal field presents important ethical challenges, particularly concerning gender bias. As AI systems are applied to areas like criminal justice, hiring decisions, and legal risk assessments, there is an increasing risk that these technologies may reinforce existing biases. The potential consequences are significant, as unchecked gender bias in AI threatens to undermine legal decisions and worsen disparities. As AI becomes more influential in legal processes, addressing and correcting gender bias is crucial to ensure that these tools are used fairly and ethically, supporting the principles of justice in an increasingly automated legal landscape. Understanding the relationship between AI, gender bias, and the law is critical for fostering a more inclusive and equitable legal system.

III.1. Anti-Discrimination Law

A central legal concern about AI is its potential to introduce or reinforce gender bias in decisions typically protected by anti-discrimination laws. AI recruitment tools, for example, have at times preferred male candidates due to biased data reflecting historical employment trends, where men were overrepresented in technical and leadership roles.

Existing legal frameworks, particularly in countries like the United States and the European Union, prohibit gender discrimination in employment under laws like Title VII of the Civil Rights Act of 1964 and the Equal Pay Act.[4] Similarly, in Mexico, gender-based labor discrimination is prohibited by laws such as the Constitution, the Federal Labor Law, and the Federal Law to Prevent and Eliminate Discrimination.[5]

[4] The Title VII of the Civil Rights Act of 1964 prohibits employment discrimination based on race, color, religion, sex, and national origin. This includes gender discrimination. And the Equal Pay Act of 1963 requires that men and women be given equal pay for equal work in the same establishment.

[5] 1) Political Constitution of the United Mexican States: Article 1: Prohibits all discrimination based on gender, among other reasons, that undermines human dignity. Article 123: Establishes that everyone has the right to dignified and socially useful work, and any discrimination in labor matters is prohibited.
2) Federal Labor Law: Article 3: States that no one may be discriminated against based on gender, marital status, age, ethnicity, sexual orientation, religion, among others. Article 164: Establishes equality of rights and opportunities between men and women in the workplace.

However, AI introduces a new layer of complexity: discrimination is often unintentional and difficult to detect. AI systems can subtly introduce biases, operating as a "black box" that obscures the decision-making process, making difficult to prove discrimination.

This raises several legal questions. Can developers or employers be held liable for gender discrimination perpetuated by an AI system? How can the bias in training data be proven, and who is responsible for it? These issues challenge traditional anti-discrimination law, which generally relies on clear evidence of intent or overt differential treatment.

III.2. Transparency and Accountability in AI Systems

Transparency is one of the main challenges in regulating AI. AI algorithms, especially those based on machine learning, are often opaque. Even developers may not fully understand how decisions are made. This lack of transparency poses significant barriers to legal accountability. If a person is discriminated against by an AI system —whether in hiring or other areas— it can be difficult for them to challenge the decision if they cannot understand why it was made.

Legal scholars have highlighted data privacy laws like the General Data Protection Regulation (GDPR) in the European Union as offering some protections against this opacity. Under Article 22 of the GDPR,[6] individuals have the right to contest decisions made solely by automated means if those decisions significantly affect them. This provision is crucial for addressing gender bias in AI, as it gives individuals a mechanism to challenge potentially biased outcomes and forces companies to explain their automated processes.

However, while the GDPR provides a legal pathway for accountability, it does not entirely resolve the issue of AI transparency. Simply requiring companies to explain their AI processes does not necessarily reveal hidden discriminatory patterns, especially when those patterns are deeply embedded in the data. A more robust legal framework may be needed, one that mandates regular audits of AI systems, particularly in sectors where discrimination can have serious consequences.

3) Federal Law to Prevent and Eliminate Discrimination: Article 4: Defines discrimination as any distinction, exclusion, or restriction aimed at nullifying or impairing the rights of individuals, including labor rights, based on gender, among other reasons.

[6] Article 22 of the GDPR states that individuals have the right not to be subject to decisions based solely on automated processing, including profiling, if such decisions have significant legal or similar effects on them. This article is particularly relevant when it comes to AI systems, as it allows individuals to challenge decisions made without human intervention, thereby offering a layer of protection against opaque AI decision-making.

IV. The Role of Regulatory Oversight and Ethical AI

The legal implications of gender bias in AI extend beyond anti-discrimination law. As AI becomes more prevalent in areas such as healthcare, finance, and criminal justice, the potential for systemic bias becomes a societal concern. This underscores the need for comprehensive regulatory frameworks to address the ethical use of AI.

The European Commission has developed Ethics Guidelines for Trustworthy AI, which emphasize fairness, transparency, and accountability. These guidelines advocate for AI systems that actively avoid reinforcing gender bias and other forms of discrimination. From a legal perspective, a pressing question is whether these guidelines should become enforceable laws. While guidelines offer a roadmap for ethical AI development, they lack the force of law. Making them mandatory would provide a stronger foundation for ensuring that AI systems are regularly evaluated for bias, and that those responsible for developing and deploying them are held accountable for any discriminatory outcomes.

V. Conclusion

Gender bias in AI poses a significant legal challenge, particularly in the context of existing anti-discrimination laws. As AI systems play a growing role in decisions affecting individual rights and opportunities, legal frameworks must adapt to address the unique challenges posed by this technology. Transparency, accountability, and fairness must be at the core of AI regulation, ensuring that AI does not perpetuate gender bias and that individuals have the ability to contest discriminatory outcomes.

A potential solution is to implement legal mandates for conducting bias audits. These independent audits of AI systems could uncover and address gender biases before they result in harm. Furthermore, developers could be obligated to show that their systems have undergone thorough fairness testing and that measures have been taken to reduce any biases identified. This strategy would be consistent with the principles of due diligence and accountability that are applied in other legal domains.

Strengthening legal protections and regulatory oversight is essential for harnessing the benefits of AI while safeguarding against its potential harms. Only through deliberate action can we create AI systems that are truly inclusive and just.

VI. References

Buolamwini, Joy y Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", en *Proceedings of the 1st Conference on Fairness, Accountability, and Transparency (PMLR)*, 2018 [en línea], <https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.

Dastin, Jeffrey, "Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women", en *Reuters*, 10 de octubre, 2018 [en línea], <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.

Eubanks, Virginia, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poo*, [s.d.], St. Martin's Press, 2018.

European Commission, *Ethics Guidelines for Trustworthy AI*, 2020 [en línea], <https://ec.europa.eu/futurium/en/ai-alliance-consultation>.

Haraway, Donna, "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective", en *Feminist Studies*, EUA, vol. 14, núm. 3, 1988.

Harding, Sandra, *Whose Science? Whose Knowledge? Thinking from Women's Lives*, [s.d.], Cornell University Press, 1991.

Intersoft Consulting, General Data Protection Regulation (GDPR), Art. 22 GDPR, Automated Individual Decision-Making, Including Profiling, 2024 [en línea], <https://gdpr-info.eu/art-22-gdpr>.

Noble, Safiya Umoja, *Algorithms of Oppression: How Search Engines Reinforce Racism*, [s.d.], New York University Press, 2018.

U.S. Department of Labor, Equal Pay Act of 1963, 2024 [en línea], <https://www.dol.gov/agencies/wb/equal-pay-act>.

U.S. Equal Employment Opportunity Commission (EEOC), Title VII of the Civil Rights Act of 1964, 2024 [en línea], <https://www.eeoc.gov/statutes/title-vii-civil-rights-act-1964>.

Wachter, Sandra, Brent Mittelstadt y Luciano Floridi, "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation", en *International Data Privacy Law*, Oxford University Press, vol. 7, núm. 2, mayo, 2017 [en línea], <https://doi.org/10.1093/idpl/ipx005>.

Wachter-Boettcher, Sara, *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech*, [s.d.], W.W. Norton & Company, 2017.